Department for Transport

TAG UNIT M1.2

Data Sources and Surveys

January 2014

Department for Transport

Transport Analysis Guidance (TAG)

https://www.gov.uk/transport-analysis-guidance-webtag

This TAG Unit is guidance for the **MODELLING PRACTITIONER**

This TAG Unit is part of the family **M1 – MODELLING PRINCIPLES**

Technical queries and comments on this TAG Unit should be referred to:

Transport Appraisal and Strategic Modelling (TASM) Division
Department for Transport
Zone 2/25 Great Minster House
33 Horseferry Road
London
SW1P 4DR
tasm@dft.gsi.gov.uk

Contents

1 Introduction

1.1 Introduction to Data Sources

1.1.1 The aim of this Unit is to identify what sources of transport data are available to practitioners developing transport models. The unit also addresses the methods used for gathering data including survey methodology.

1.1.2 Data collection is necessary in order to build a transport model and to inform the numerical parameters used in the model, many of which cannot be observed directly, and therefore need to be estimated using a sample of data (**model calibration**). Models should be compared against independent observed data for the current or a recent year to demonstrate that the model is of sufficiently good quality (**model validation**) to form the basis from which to calculate future forecasts.

1.1.3 One of the main constraints on data collection is the cost of conducting transport surveys. This cost can be reduced by minimising the need to collect new data. In order to reduce the costs of building a transport model, the practitioner should make use of all relevant and available transport datasets.

1.1.4 This unit lists the major available datasets and describes the data which are contained within them. It addresses transport data which are relevant to the development of demand models as well as supply or assignment models.

1.1.5 This unit also contains advice on conducting study specific surveys. It addresses the requirements of each type of survey and discusses the standards to be expected with each survey type including survey accuracy.

1.2 Structure of this Unit

1.2.1 After this introductory section, this unit has been divided into two main sections. Section 2 contains guidance on:

- available datasets for demand modelling including TEMPRO / NTEM data;
- highway traffic datasets;
- Public Transport (PT) passenger data; and
- network data.

1.2.2 Section 3 contains guidance on conducting surveys for specific studies and includes:

- data specification and requirements for demand models;
- conducting highway traffic surveys and the standards and tolerances required for the data; and
- PT surveys and the methods to be used for gathering these data.

2 Existing Data Sources

2.1 Introduction to Existing Sources

2.1.1 This section identifies existing sources of transport data held by the Department and its agencies, what these existing datasets contain and where those datasets currently reside.

2.1.2 Taking advantage of existing transport data can be of vital importance in the specification and implementation stages of model development as referred to in the Guidance for the Technical Project Manager, and will reduce costs.

2.2 Planning and Demand Data

2.2.1 Multi-modal studies should be undertaken using a transport model, with demand changes controlled by exogenously defined planning data (possibly with some feedback to land-use).

2.2.2 It is highly desirable that the planning data used should at some level be consistent with the DfT NTEM (National Trip End Model) projections, so as to ensure that different regions do not engage in competitive bidding by assuming local growth rates which are implausible when summed across the whole economy. NTEM projections can be accessed using TEMPRO (Trip End Model Presentation Program).

2.2.3 TEMPRO itself is a program used to distribute the results from the NTEM forecasts. The existing NTEM system works mainly at the local authority district level. Forecasts of population, households, workforce, and jobs for each district over 30 years have been prepared by the Department, after consultation. These forecasts are used in a series of models which forecast population, employment, car ownership, trip ends and traffic growth by district. The NTEM dataset can be viewed using the TEMPRO software. Both are available free of charge on the TEMPRO website: https://www.gov.uk/government/organisations/department-for-transport/series/tempro

2.2.4 TEMPRO is designed to allow detailed analysis of pre-processed trip-end, journey mileage, car ownership and population/workforce planning data from the NTEM. The pre-processed data are themselves the output from a series of models developed and run by the Department. TEMPRO can also be used to provide summaries of traffic growth using data from the National Transport Model (NTM).

2.2.5 The current version is multi-modal, providing data on trips on foot, by bicycle, motor vehicle (both as a driver and passenger) by rail and by bus. Users should note, however, that TEMPRO trip ends by mode are based on average rates over a wide area and do not necessarily take into account the accessibility of each zone by each transport mode.

2.2.6 The use of a land-use / transport interaction (LUTI) model to provide the 'planning data' inputs in interaction with the transport model would obviously raise a different and more specialised set of data requirements which we do not attempt to consider here.

2.3 Traffic Data

2.3.1 The Department has an extensive database of traffic count information available. The following geographical website, www.gov.uk/government/organisations/department-for-transport/series/road-traffic-statistics, provides Annual Average Daily Flow (AADF) and traffic data for every junction-to-junction link on the 'A' road and motorway network in Great Britain. It enables the user to search, view and download these data for each count point on the major road network, and view regional summaries.

2.3.2 An interactive map on the website provides a mapped background to identify traffic flows in specific areas of the country. Data from a sample of points on the minor road network are also available to download from the website.

2.3.3 Complete datasets are also available from the website http://data.gov.uk/. These include road traffic counts, local authority traffic estimates and traffic, speed and congestion data. Before accessing the data users will need to create an account at http://data.gov.uk/user/register.

2.3.4 Journey time information derived from a variety of sources including Global Positioning System (GPS) and Automated Number Plate Recognition (ANPR) data can be accessed through the Department. These data cover the Great Britain road network and contain journey time information disaggregated into 15 minute segments.

2.3.5 The primary use for these data by the Department is to monitor congestion levels and the reliability of journey times on the road network. However, the data may be of use to model developers to validate model network speeds. Access to these data can be given to practitioners working on behalf of the Department or any of its agencies by contacting the Congestion Statistics team in the Department.

2.3.6 Highways England also has a large amount of traffic data available and it is expected that full use will be made of this information. A large source of highway traffic information is the Highways Agency Traffic Information System (HATRIS). HATRIS consists of two traffic databases including the Journey Time Database (JTDB) and the Traffic Flow Data System (TRADS).

2.3.7 The TRADS dataset is a central collection and reporting point for 15 minute and hourly based traffic flow information. Traffic flow information from 1,500 roadside inductive loops on the strategic highway network can be obtained from this database. The JTDB is a processing and reporting system holding all journey time data fed in via a number of sources including roadside cameras and fleet GPS.

2.3.8 Traffic flow and journey time information from HATRIS can be obtained from the website http://www.hatris.co.uk/ after first registering and obtaining a secure username and password to log onto the HATRIS website.

2.3.9 In some cases, traffic data will also be available from local authorities. It is advised that users contact the authority in their area of concern to inquire about the availability and extent of traffic data.

2.4 Public Transport Data

2.4.1 As an input to PT models, it is possible to obtain rail passenger matrices from the LENNON rail ticket information database. As an initial stage, the demand estimate derived from the processing of LENNON data is a station to station trip matrix.

2.4.2 It should be noted that LENNON information has a number of limitations:

- it does not cover tickets bought for travel on the major light rail or underground systems of the UK, including the London Underground;
- LENNON cannot assign a point to point journey where a multi-modal ticket is bought (e.g. rail/bus/LRT travelcards);
- LENNON does take account of travel that takes place without a ticket (e.g. free concessionary travel, PTE rail only travelcards) but has limitations with flat fare concessionary travel;
- output LENNON information covers all days of the week (a limitation brought about by the lack of recorded data concerning the return leg of return ticket purchases, and the timing of use of season tickets);
- LENNON information is not segmented by purpose or car ownership; and
- LENNON data are not collected in production / attraction (P/A) format.

These limitations will need to be considered carefully before a decision is taken to use this data set.

2.4.3 LENNON data are owned by the Association of Train Operating Companies (ATOC) and, although the Department has access to the LENNON database, the Department can only share these data with third parties working on its behalf. Any practitioners not conducting work on behalf of the Department and requiring access to LENNON data should first contact the ATOC.

2.4.4 The National Rail Travel Survey (NRTS) is a survey of passenger trips on the national rail system in Great Britain on weekdays outside school holidays. It was carried out for the London Area as part of the London Area Transport Survey (LATS) in 2001 and throughout the rest of Great Britain in 2004 and 2005. This dataset includes rail passenger information disaggregated by origin / destination, ticket type, trip purpose, time of travel and demographic information. Aggregate and record level data are available from the Rail Statistics team in the Department, although personal data (such as origin post-codes) are only available for those working on behalf of the Department.

2.4.5 Information on rail passenger counts is available from the Department. These data contain information on passenger loads and capacity provision on individual train services. The Department also publishes aggregate statistics showing arrivals and departures at a number of cities and on routes into central London in one hour time bands throughout the day on a "typical" autumn weekday and is available at https://www.gov.uk/government/publications/. The information for individual services can be shared with third parties working on behalf of the Department, other users should contact the individual train operating companies.

2.4.6 Electronic ticket machine (ETM) data provide information on all journeys rather than a sample. ETM data provide time of travel and can be obtained over long periods of time, thereby avoiding day-to-day variations. They can also relate to the network of services as a whole. However, ETM data will only provide trip records in terms of fare stages at which passengers board or alight, and fare stages may differ between different operators. In areas where the use of travel cards, concessions and other pre-paid tickets are prevalent, ETM data may provide a less accurate picture of passenger movement. Practitioners should contact bus operators in their area of interest to enquire about the availability of ETM data.

2.5 Other Sources of Data

Census Data

2.5.1 Further information on commuters journey to work and their movements is also available from the 2001 and 2011 census journey to work data.

2.5.2 The census data should be used with caution, for two main reasons. First, this data source only includes trips between home and work and does not provide any information about trips for other purposes. Secondly, the correspondence between the trips recorded in the census and those that actually take place is not direct. The census records 'usual journey to work' but the average person doesn't complete their usual journey to work every day due to sickness, work leave, etc. At the aggregate level therefore the census data over-estimates journey to work trips. Nevertheless, the census journey to work data may be of use in developing transport models (e.g. providing useful information on trip distributions) as long as it can be supplemented by data from other sources.

2.5.3 Census journey to work data can be downloaded from the Government data website http://data.gov.uk/ and also at http://www.nomisweb.co.uk/.

National Travel Survey (NTS) Data

2.5.4 The NTS is the primary source of data on personal travel patterns in Great Britain. The NTS is an established household survey which has been running continuously since 1988. The survey is primarily designed to track the long-term development of trends in travel, although, short-term trends can also be detected.

2.5.5 NTS data are collected via two main sources – interviews with people in their homes and a travel diary that they keep for a week to record all their trips. The NTS covers travel by all age groups, including children. Typically the sample size every year since 2002 has been 8,000 households and 20,000 individuals, although this can vary year by year.

2.5.6 The NTS results and technical information can be downloaded at the following website: https://www.gov.uk/government/organisations/department-for-transport/series/national-travel-survey-statistics. This website also contains contact information for the NTS team in the Department.

Aviation Data

2.5.7 The Civil Aviation Authority (CAA) publishes a range of information on its website about the usage of airports around the UK. This includes summary results from the CAA Passenger Survey which contains information about the geographic origin or destination of air passengers, their purpose of travel and passenger characteristics. This survey is completed on an annual basis at Birmingham, Gatwick, Heathrow, Manchester and Stansted with a selection of other airports covered each year. Access to the full dataset is available from the CAA (contact details on their website, http://www.caa.co.uk) for a fee.

2.6 Network Data

2.6.1 A significant number of national highway network datasets are now available and they can be of use in transport model development. They range in level of detail from:

- detailed road centre line representations showing all highways and details of curvature; through to
- skeletal representations of major roads used for 'journey planner' packages.

2.6.2 The detailed datasets are not generally suitable for direct incorporation into models, as they contain a vast amount of unnecessary detail that would complicate the model and increase memory requirements and processing times. Simplification of some of the data contained within network datasets is therefore required. However, practitioners should consider retaining some geometric data that might be required for environmental analyses, especially where the network is stored in a GIS (Geographic Information Systems) package.

2.6.3 Care should also be taken when interpreting data from freely available network datasets and journey planners. For instance, the journey planner systems usually contain information on link travel times that are generally an inter-peak average rather than representative of any particular peak time period unless specifically stated.

2.6.4 The Ordnance Survey MasterMap Integrated Transport Network (ITN) is a useful resource for verifying and building modelled networks. In conjunction with GIS packages, it can also be used as a background for modelled networks to help with comprehension of modelled representation of networks and for environmental analyses. The ITN can be downloaded from the OS OpenData website at www.ordnancesurvey.co.uk/opendata.

2.6.5 The Department currently maintains the major roads database using information from the OS ITN. The major roads database contains a series of links which are unique sections of road which make up the entirety of the major roads network in Great Britain. This database contains various fields of information about these links including link length, available traffic count information and other physical characteristics for all links greater in length than 300 metres.

2.6.6 Other important sources of data for road network modelling include:

- junction plans held by local authorities and Highways England;
- traffic signal timing and phasing data held by local authorities or traffic control units;
- site surveys involving observation of traffic and parking behaviour;
- aerial photographs; and

- online mapping and satellite imaging such as Google Maps and StreetMap.

2.6.7 The National Public Transport Access Node (NaPTAN) database is a UK nationwide system for uniquely identifying all the points of access to public transport in the UK. It is a core component of the UK national transport information infrastructure and is used by a number of other UK standards and information systems. Every UK public transport node including rail stations, coach termini, airports, ferry terminals and bus stops are allocated at least one identifier.

2.6.8 The NaPTAN database system is a UK national standard sponsored by the Department and supports both the public registration of bus timetables by the Vehicle and Operator Services Agency (VOSA) and the data collection for the Transport Direct Portal. Data from NaPTAN may be obtained at the website http://www.dft.gov.uk/naptan/

2.6.9 The National Public Transport Data Repository (NPTDR) database contains a snapshot of every public transport service in Great Britain for a selected week in October each year. Data is currently available from October 2004 onwards. The dataset is compiled with information from many sources including public transport information from each of the Traveline regions, coach services from the national coach services database and rail information from the Association of Train Operating Companies.

2.6.10 NPTDR can also be used in the production of accessibility indicators via software tools such as Accession. Further information on the NPTDR database, and the data themselves, may be found at the following website http://data.gov.uk/dataset/nptdr.

2.6.11 The nature of the deregulated bus market, with service changes permitted at short notice, means that precise coding of public transport supply from published information is often not possible. However, bus networks are now more stable than in the years immediately following deregulation. Changes that do occur are often limited to minor adjustments to service frequencies and stopping points, often below the level of significance that would require a major adjustment to model networks. Within London, the absence of deregulation means that networks are relatively stable and published service information is generally up to date. Data for rail network modelling are easier to obtain than for bus and the level of detail available is generally greater. A rail timetable gives precise station to station travel times for each departure, whereas for buses timetable timing points can be several stops apart.

3 Study Specific Surveys

3.1 Introduction to Study Specific Surveys

3.1.1 This section of the unit provides guidance on conducting bespoke transport surveys for individual study requirements.

3.1.2 Data collection is often the largest cost of building a transport model. This cost can be reduced by minimising the quantity of new surveys. The best way to minimise data collection requirements is by taking data needs into account in the design of the model. Most models split the spatial area to be studied into sectors and are further sub-divided into zones. Data collection can be reduced by careful design of the sectors and zones. For guidance on zonal design, see TAG Unit M3.1 - Highway Assignment Modelling.

3.1.3 Methods for data collection can generally be divided into three categories, listed in ascending order of cost:

- non-intrusive automated methods (for example automated traffic counts; emergent data sources including data from electronic devices may fall into this category);
- non-intrusive manual methods (for example, traffic counts or manual turning count surveys collected by an observer, or observed journey time surveys); and

- intrusive methods (interviews with potential customers, which include road side interviews and travel diary surveys).

3.1.4 Where there is an option, therefore, the top priority is often to minimise the reliance on intrusive data collection methods, as this is likely to give the greatest cost saving, although some travel information (e.g. trip purpose) cannot normally be gathered without some form of intrusive surveys.

3.2 Demand Data Surveys

3.2.1 The advice below lists the type and sources of data required for variable demand modelling, beyond the data used for assignment modelling. In most cases only a sub-set of these data sources will be needed for a given model; less detailed models will require less data, and may base some of their segmentation on assumptions or transfer of appropriate data from other models. As a minimum, however, any multi-stage demand model will require a database which provides, in the base year for each origin zone, the total number of trips:

- by car availability;
- to each destination zone;
- for each of several purposes;
- by each of the modes modelled; and
- within each modelled time period.

3.2.2 Table 1 which follows list the main data requirements for variable demand modelling.

Table 1 Demand Model Data Requirements

Data Required	For use in	Sources	Notes
Population	Trip generation	NTEM, Census	From wards or Census Output Areas amalgamated to zones: advisable to categorise by sex and adult/child
Households	Trip generation	NTEM, Census	From wards or Census Output Areas amalgamated to zones
Car ownership	Trip generation, Modal split, Distribution	NTEM/NATCOP, Census, Local Household Travel Survey, Household Expenditure Survey	Averaged across zones, or might be estimated from other socio-economic data
Socio-Economic Group	Trip generation	Census, Local Household Travel Survey, Household Expenditure Survey	Not often used except in distinguishing workers from unemployed and retired, but some models categorise work travel by broad SEG groups (e.g. blue/white collar) and land-use models also require considerable SEG data.
Land-use data	Trip generation, Distribution	Census and Special Workplace Statistics, NTEM for trip ends by purpose, Local Authority planning data for employees, retail and commercial floorspace by zone.	Employees and retail space by zone enable calculation of zonal trip totals for doubly-constrained (to/from work) trip distribution, and attractions for optional trips. Some LUTI models need considerable detail for Social Economic Group (SEG), employment and floorspace data.
Car availability	Modal split, Distribution	Local Household Travel Survey, National Travel Survey (NTS)	Can be estimated from household composition and license holding
Licence holding	Trip generation, Modal split	Local Household Travel Survey, Household Expenditure Survey, National Travel Survey	Categorisation by number of licences and number of household cars indicates level of car availability
Travel to zonal destinations by: Purpose, Mode and Time of Day(Period)	Distribution, Modal split, Time-period choice	Roadside Interviews (RSIs), Local Household Travel Survey, journey to work from Census, NTEM, NTS	Full detail only available from specialised local surveys, by interview or questionnaire. NTS local samples are small, but values might be adjusted from wider NTS (e.g. split of purpose by time of day). Census journey to work may be extrapolated to other purposes via NTEM. Mode might include active modes and distinguish between car driver and passenger.

Trip lengths	Distribution, Modal split	Local Household Travel Survey, National Travel Survey, journey to work Census, RSI surveys	Not used directly since distance is specified by zonal structure, but should be used in validation
Vehicle operating costs	All responses	Values from TAG Unit A1.4, Values of Time and Operating Costs	Perceived money costs of a vehicle journey are less than true average cost, and are assumed to be different for business and private travel.
Vehicle occupancies	All responses – for estimating costs	RSIs, local Household Travel Survey, NTS	By purpose; often assumed from NTS or NTEM or TAG Unit A1.4
Values of time	All responses	TAG Unit A1.4, Values of Time and Operating Costs	Differ by purpose, and updated as appropriate. In principle, different values for behavioural modelling and standard values for appraisal

3.2.3 The sources in Table 1 are not independent. The National Trip End Model (NTEM) offers a valuable source of most of the data required to predict trip ends, both trip productions based on household characteristics and trip attractions based on employment etc, as well as car availability forecasts.

3.2.4 In many cases NTEM may be sufficient, and there will be no need to go back to the original sources. In other cases the NTEM zones may be too large, so that the NTEM data may have to be adjusted in the light of local knowledge or data specific to the study zones obtained from more original sources. Effort can be saved if there has been a suitable local transport survey conducted, or if there is a local validated transport model from which the required data can be extracted.

Household Interview Surveys

3.2.5 Household surveys provide the most complete picture of travel by residents of a study area, including walking and cycling. Outputs from these surveys can be segmented by the key variables of household type, person type, trip purpose, mode and time period. However, building of trip matrices directly from household surveys is not generally practical on the grounds of cost - the method is inherently expensive per person trip recorded. The primary application of this type of dataset is therefore in the segmentation of demand data collected from other sources, and for use in creating local car ownership and trip end models where the national models are not thought to be appropriate.

3.2.6 Household interview survey sample sizes are rarely sufficiently large to provide acceptably accurate estimates of trips between pairs of zones. Therefore, data from this source should not generally be used directly in the creation of trip matrices at a zonal level. However, household interview surveys are a rich data source, in the sense that actual trip-making behaviour by all modes can be linked to the characteristics of the household and travellers. This data source is useful, therefore, for demand model estimation.

Revealed Preference and Stated Preference Surveys

3.2.7 Revealed Preference (RP) refers to observations of actual behaviour, for example the mode choices that decision-makers currently make or made in the past. RP data are inherently more credible than Stated Preference (SP) data and their use, if only partially, will strengthen the credibility of demand forecasts in the appraisal framework.

3.2.8 SP refers to observations of hypothetical behaviour under controlled experimental conditions. A scheme that introduces a new mode, for example, would imply a need for SP analysis, since RP data are by definition unavailable for such a context. Developing a bespoke mode choice model, therefore, often requires new SP surveys and analysis.

3.2.9 RP data can be obtained from SP respondents, from postcard surveys (an under-used and relatively inexpensive approach), from home or phone interviews, travel diaries, as well as from the National Travel Survey and Census.

3.2.10 The collection of RP data is not without problems. There are often large biases in respondents' self reported data, underestimating the costs of their chosen mode and overestimating the costs of alternative modes. To overcome these problems it is sometimes necessary to use explanatory variables from network models and published timetable data. Even where respondents' reported data are modelled, there is often a considerable amount of missing data which needs to be collated.

3.2.11 For more information on SP and RP surveys see Supplementary Guidance - Bespoke Mode Choice Models.

3.3 Highway Surveys

3.3.1 This section covers the typical highway traffic surveys which are carried out. Highway traffic surveys are generally carried out for three purposes: matrix creation, model calibration and validation.

3.3.2 Calibration and validation data are of two kinds: traffic counts, and journey times while RSI surveys are commonly used in the matrix creation process.

3.3.3 **Traffic counts** are required for:

- expanding new roadside interviews;
- re-expanding old roadside interviews;
- calibrating trip matrices by means of matrix estimation; and
- validating the model.

3.3.4 **Journey times** are required for:

- calibrating cruise speeds (speeds between junction queues);
- identifying where delays occur at junctions; and
- validating the model.

3.3.5 Traffic counts may be obtained by automatic means (Automatic Traffic Counts, ATCs) or manually (Manual Classified Counts, MCCs). Journey times may be obtained by Moving Car Observer (MCO) surveys or from commercial sources of tracked vehicle data, or camera observations from Automatic Number-Plate Recognition systems (ANPR) or from such traffic databases listed in section 2 of this unit. In selecting the appropriate type of count and source of journey times, these factors need to be considered:

- the accuracy of the data;
- the choice of survey locations;
- the need for information by vehicle type; and
- a recognition of the costs of these data.

3.3.6 Surveys should be carried out during a 'neutral', or representative, month avoiding main and local holiday periods, local school holidays and half terms, and other abnormal traffic periods. National experience is that the following Monday to Thursdays can be neutral:

- late March and April – excluding the weeks before and after Easter;
- May - excluding the Thursday before and all of the week of each Bank Holiday;
- June;
- September – excluding school holidays or return to school weeks;
- all of October; and
- all of November – provided adequate lighting is available.

This requirement often dictates the timescale of the appraisal. Data processing may also add substantially to the study timescale.

3.3.7 In addition, if existing data are to be reused, ample time must be allowed for them to be identified, obtained from their current custodian, reprocessed as necessary, and checked for consistency and validity. Further delays may be incurred if these checks reveal that the data cannot be used.

Traffic Count Surveys

3.3.8 Manual classified counts (MCC) are required to break down traffic flows by vehicle type. This information is particularly important in an urban area, where the mixture of vehicle types may vary significantly by direction as well as at different times of day. Classified counts are required at every roadside interview site, and on minor parallel roads not included in the interview programme, to expand the interview sample to the total traffic flow in the corridor as a whole (see Paragraph 3.3.25). Counts should be carried out in both directions on the survey day, even if interviewing is only in one direction, and should extend over all model periods. If automatic counts indicate that traffic flows at a roadside interview site were influenced by the presence of the interview survey, further manual classified counts should be made on a different day. If necessary, these alternative counts can then be used to expand the interview data to a more representative traffic flow. The vehicle classification used should correspond with that used in the interview survey itself, and this in turn should be compatible with the vehicle types represented in the traffic model.

3.3.9 Turning counts at road junctions are required for the calibration and validation of junctions in a congested assignment model. Turning counts should be carried out at all junctions within the model area that are likely to have a significant impact on journey times or delays and at junctions that are particularly significant in route choice (i.e. locations where alternative routes for critical movements may merge/diverge). In urban areas, there will often be a need to collect more turning count data than for an inter-urban model, because of the greater number of junctions that generally need to be validated.

3.3.10 Turning counts are carried out in the same manner as manual classified counts on links, except that more enumerators are generally required. They must cover the whole of each peak period, but need only cover representative parts of other time periods, depending on the time periods being modelled. Where an inter-peak model is representing an average inter-peak hour, a 12 hour period covering the two peaks and the inter-peak would be required. The vehicle classification used may be simpler than the one used for link surveys, provided that it is again compatible with the model classifications. For more complex or larger junctions, video or ANPR surveying methods may need to be employed to fully cover all turning movements at the junction.

3.3.11 Automatic traffic counts (ATC) and carrying them out is an operation requiring a substantial investment in instruments, ancillary equipment, transport, data handling systems and staff time. The volume of data that can be collected is considerable and can reveal longer term traffic volume trends, but the effort expended could prove fruitless if any one of the constituent processes involved in collecting and processing the data is deficient.

3.3.12 ATCs are used for all the purposes mentioned in paragraphs 3.3.8 to 3.3.11 and, in addition, can be used to monitor traffic flows and to provide information about the relationship between survey day traffic and longer term flow levels. Automatic counts can also be used to provide information about local 12-, 16-, 18- and 24-hour flow ratios, and daily and seasonal traffic variations, all of which are required to estimate average daily traffic flows from shorter period data.

3.3.13 Little definitive work has been published concerning the accuracy of traffic counts by automatic traffic counters. Experience suggests that the errors are machine and (particularly) installation dependent. TRRL Supplementary Report SR 514 "Estimation of annual traffic flow from short period traffic counts" contains some useful results on the efficiency and accuracy of annual estimates from short period counts. For longer term counts, the frequency and diligence of the station monitoring and servicing will be crucial.

3.3.14 Good quality counts are required for validation (as well as for calibration). The notion of using good quality counts for calibration and poorer quality counts for validation, or vice versa, should **not** be considered.

Journey Time Surveys

3.3.15 Journey time measurements are an essential part of assignment model validation for most urban traffic appraisals, since the majority of scheme benefits tend to be related to journey time savings. Comparison of observed and modelled journey times gives a measure of the appropriateness of the speed-flow relationships for a capacity restrained assignment, as well as the junction delay calculations for a congested assignment model. Journey time surveys may also be used to identify junctions which exhibit high levels of delay and that need to be modelled in detail.

3.3.16 For validation of journey times along selected routes, MCO data should be used where available (note that some MCO surveys will be necessary in any event in order to verify data from other sources, such as tracked vehicle data). For other routes, data from other sources (such as HATRIS) may be used, if proven to be of sufficient accuracy (see paragraph 3.3.20) and not materially biased. These comparisons will show how well total link times are modelled.

3.3.17 For general purpose models, the routes for the validation of journey times should cover as wide a range of route types as possible and cover the fully modelled area in the model as evenly as possible. For models developed for the appraisal of specific interventions, routes should include those on which it is expected traffic will be affected by the scheme, as well as covering the route including the proposed scheme itself, if appropriate.

3.3.18 The validation routes should be neither excessively long (greater than 15 km) nor excessively short (less than 3 km). Routes should not take longer to travel than the modelled time periods (although, a few minutes longer is unlikely to be problematic). Start times should be staggered, particularly if runs are undertaken on the same day. For models of actual peak hours, journey time routes ought to be no longer than about 40 minutes to allow some staggering of start times.

3.3.19 During the survey, the total travel time should be recorded separately for each road section between major junctions and, because junction delays form an important part of travel time, a separate note should be made of the delay time at each junction. Ideally, delay should be assumed to start once instantaneous speed falls below a chosen speed, say 15 kph (10 mph). Journey time runs, in both directions and in each model time period, should be made over a period of several days. Variations in travel times during peak periods should be taken into account by staggering start times, to represent fairly conditions over the time period as a whole.

3.3.20 In the case of journey times for all vehicles combined, sufficient MCO runs should be undertaken so that the 95% confidence level of the mean of the observations is ± 10% or less over a route as a whole. Four initial journey time runs (preferably each on a different day) should be made for each route, direction and model time period, and the results used to assess the variability of journey times in each case. Further runs must be made for those routes, directions or time periods in which the variability falls outside the acceptable range. In urban areas, where journey times can sometimes be erratic, this may lead to a large number of runs being required. If a satisfactory level of consistency has not been achieved after 12 runs, the results should be accepted and a special note made in the survey documentation. The accuracy of journey time data derived from other sources, such as tracked vehicle data or HATRIS, should also be determined and taken into account when making comparisons between modelled and observed times. Further detail on the accuracy of local journey time measurements is available in the DMRB manual 13.1.5.11.

3.3.21 Using separate speed/flow relationships for light and heavy vehicles, journey times applicable to these two classes of vehicle would be desirable, but not essential. For light vehicles, appropriate journey times may be obtained by restricting the Moving Car Observer method to light vehicles. For heavy vehicles, other techniques, such as video registration plate surveys may be required.

Road Side Interview (RSI) Surveys

3.3.22 Detailed advice on conducting and the collection of roadside interview data can be found in the Department’s DMRB volume 12, section 1 part 1, chapter 6 and in particular in DMRB volume 5, section 1, Part 4, TA11/09 “Traffic Surveys by Roadside Interview”. The RSI survey process is disruptive and has inherent but manageable risks. Safety to the staff involved and the general travelling public can be mitigated and data integrity improved by following the advice in those documents.

3.3.23 Addresses should be coded either to Ordnance Survey Grid References (at least six digits) or to Post Codes. Adopting these minimal standards will ensure that data can be re-used by others at a later date.

3.3.24 Analysts should note that roadside interviews cannot be conducted on motorways. In addition, it may be impractical to conduct roadside interviews on some all purpose roads, especially where they carry high levels of traffic. Relocation of interview sites may address these problems - for example, interviewing on motorway entry slip roads may provide an alternative to interviewing on motorways.

3.3.25 RSI surveys should always be associated with a MCC survey carried out on the day of the RSI survey and a minimum two-week ATC survey. New roadside interview records should be expanded by vehicle type from the MCC on the day of survey. The surveys should be scheduled so as to minimise potential for diversions to avoid the interview site, thereby making the MCC on the day of survey atypical.

3.3.26 A number of other adjustments are required, as follows:

- conversion from the day of survey to an average weekday;
- adjustment to account for diversions on the day that the interviews were conducted;
- adjustment to account for day to day variability; and
- conversion from the month of survey to an average neutral month.

3.3.27 Controlling to two-week ATCs will address the first and second of these adjustments and the third to an extent. The fourth adjustment will require factors to be developed from long-term ATCs. It should be noted that the application of each of these factors will add further uncertainty and widen the confidence intervals.

3.3.28 Traffic counts are also required to re-expand old roadside interview records. In principle, single day MCCs and two-week ATCs are required. However, bearing in mind the confidence intervals for heavy vehicles, it may not always be possible to justify the additional expense of the MCCs. If MCCs are not taken, the re-expansion will have to be done on the basis of the two-week ATC alone.

3.3.29 Guidance is also available on calculating required sample sizes when conducting RSI surveys. DMRB volume 12, section 1 part 1, Appendix D13 contains advice regarding the question of estimating the sample size needed to give results to the level of accuracy needed.

Data Accuracy

3.3.30 When an estimate of traffic flow has been made, it is desirable to know not only the estimated value but also how reliable this estimate is. A convenient way of expressing the precision is to state limits which, with a given probability (usually 95%) include the true value. It is then possible to state, for example, that the true value is unlikely to exceed some upper limit, or to be less than a lower limit, or to lie outside a pair of limits. This information may be more important than the estimate itself. These limits are known as “confidence limits”, ie they are limits within which it can be stated, with a given degree of confidence, that the true value lies.

3.3.31 The adoption of a particular confidence interval implies a decision concerning the accuracy required of the information presented: for example, a 95% confidence interval accepts the chance that the true value will lie outside the given limits only 5 times in 100 occurrences. The greater the accuracy demanded of the data the wider the confidence interval will be.

3.3.32 The following 95% confidence intervals for traffic counts should be assumed:

- Automatic Traffic Counts: total vehicles: ± 5%;
- Manual Classified Counts: total vehicles: ± 10%;

  Cars and taxis: ± 10%;

  Light goods vehicles: ± 24%;

  Other goods vehicles: ± 28%;

  All goods vehicles: ± 18%

3.3.33 The ATC confidence intervals relate to counters with tube vehicle detectors; counters with inductive loop may achieve greater levels of accuracy. The accuracy of radar counters is less certain but may be assumed to be the same as that of tube counters.

3.3.34 Splits between light and heavy vehicles obtained from ATCs on the basis of a 5.2m vehicle length have been shown to be subject to wide margins of error and should not be relied upon. The Highways England HATRIS database is based on a 6.6m split, that is deemed sufficiently accurate.

3.3.35 It is normal practice for MCCs to be carried out on a single day but ATCs should be conducted for at least two full weeks. ATCs carried out for two-weeks or longer will capture some day to day variability. The confidence intervals for counts will be narrower than those listed above if more observations are carried out (for example an MCC carried out on two days), but will be wider for periods shorter than one day (for example, the morning peak).

3.3.36 Turning movement counts at junctions are normally single day MCCs (or video surveys/ ANPR at more complex junctions). Their usefulness would be increased if they were to be supplemented by two-week ATCs taken for either individual turns or on the junction entry and exit arms.

3.3.37 All data should be checked to identify and remove any that might have been affected by unusual events. Where data quality is suspect, the data should be investigated thoroughly and, if necessary, rejected.

Factoring Traffic Data

3.3.38 This section deals with the factoring of traffic count data from one base to another. Every factor has an associated reliability and the result of factoring is always to increase the confidence interval of the result. Factoring should therefore be kept to a minimum and the factor with the lowest coefficient of variation should always be chosen where a choice of factors is available.

3.3.39 National factors are derived from databases which are usually larger than those a local study can generate. Whilst it is possible to derive factors locally, a full understanding of the accuracy of such factors is desirable to ensure that local conditions are indeed significantly different from the national average. This will involve not only an estimate of the coefficient of variation of the locally derived factors, but also an examination to determine whether the local factor is significantly different from the national one.

3.3.40 On some occasions more than one factor will have to be cumulatively applied in stages. DMRB volume 12, section 1, part 1, Appendix D13 gives the formulae and contains several worked examples on how this can be done and the resulting reliability of the combined results calculated.

3.3.41 The DMRB Appendix listed above also deals with the factors used to obtain an estimate of (usually) AADT from a short period count (normally 12 hours). The calculations are carried out in two stages. First the short period count is factored up to 16 hours, and secondly the 16 hour estimate is factored up to AADT. With each factor there is an associated coefficient of variation. This enables the overall accuracy of the AADT estimate to be calculated.

3.4 Public Transport Surveys

3.4.1 Many major public transport schemes in areas of established development will derive a major part of their patronage by extraction from existing public transport services. Sound information about the patterns of movements served by existing services can therefore be very important for the robustness of the appraisal of the proposed scheme.

3.4.2 As explained in TAG Unit M3.2 - Public Transport Assignment Modelling, an essential part of the development of a model for the appraisal of all major public transport schemes, except a bus priority strategy, will be a well-validated public transport passenger assignment model. A well-validated model requires a realistic representation of the network of services and good quality trip information. In most cases, the base year trip matrices should be developed from observations of travel movements with the minimum of trip synthesis, certainly in the corridors most directly affected by the proposed scheme. The following sub-sections provide advice on the types of survey that may be used to obtain data on movements by public transport passengers.

Passenger Counts

3.4.3 The numbers of passengers using a public transport system are required for:

- expanding interview samples;
- use as constraints in matrix estimation; and
- validation of trip matrices and assignments.

3.4.4 The counts required for expanding interview samples will depend on the method used to survey passenger movements.

3.4.5 Depending on passenger flows, passenger counts can be combined with on-board surveys. Often, on board face-to-face surveyors will be able to count the numbers of boarders and alighters at each stop before selecting respondents for interview. However, they will not be able to do so on crowded vehicles nor where each individual interview is likely to over-run the time between stops.

3.4.6 For multiple door vehicles such as trains, there may be substantial differences in passenger numbers counted at different sets of doors. Doors near station entrances will have a higher number of boarders than more remote doors. Because of this, an accurate level of count detail will only be available if counts are conducted on each set of doors.

3.4.7 Estimates of the numbers of passengers on public transport vehicles made by observers standing at the roadside will **not**, generally, be sufficiently accurate for any of the purposes listed in paragraph 3.4.3.

Movement Surveys

3.4.8 The information required about passenger movements for an assignment model is:

- origin and destination address;
- access mode to public transport, including any costs;
- time of travel; and
- type of ticket.

3.4.9 Other information, such as trip purpose and car availability for the journey, will usually be required for the development of the demand model. While other sources, such as a household interview survey, may provide this information, it will normally be cost-effective to collect this additional information in passenger movement surveys.

3.4.10 The main sources of information about passenger movements are as follows:

- interviews with passengers, which may be by means of face-to-face interviews with passengers, either on-board the public transport vehicles or as they wait to board at stops and stations, or by means of self-completion questionnaires; and
- electronic ticket machine data.

3.4.11 The relative merits of these sources in providing the required information are as follows.

Face-to-face on-board surveys

3.4.12 Face-to-face interviews with passengers on-board, providing the sample size is adequate, should provide the best quality data. The sample should be selected by the interviewer rather than being self-selected and it is easier to ensure that all the required information is provided. This method will enable origin and destination addresses to be provided, along with all the other information that may be required, such as access mode (and trip purpose and car availability, if required). A sample of interviews should be obtained on each service in the modelled area, thereby enabling a complete picture of travel to be obtained (once the sample data have been expanded to passenger counts).

3.4.13 In face-to-face interviews, the data can be collected from illiterate and, in the presence of guardians, juvenile travellers, which is not the case with self-completion surveys. In addition, non-response biases are reduced compared with self-completion surveys. The method also allows the surveyor to probe for particular information so that exact details of origin and destination can be discovered.

3.4.14 One of the biases that can arise from face-to-face surveys relates to length of journey – the longer the journey, the more likely that a passenger is to be sampled. To counter this bias, those travellers who are making only short hop journeys should have an equal chance of being sampled as those making longer journeys. To do this, the surveyor should select respondents by choosing a random passenger from those boarding at each stop.

3.4.15 Because more passengers will board at each stop during peak hours than inter-peak, surveyors will each be able to interview a smaller proportion of travellers during busy periods compared with inter-peak hours. This imbalance can be addressed by conducting an increased number of survey shifts during peak hours and by weighting data according to passenger counts. The important consideration is that the sample size during both the peak and inter-peak periods should be sufficiently large to support the creation of acceptably reliable trip matrices for these periods.

3.4.16 Consideration must be given to the choice of bus and public transport routes to be sampled. Ideally, each public transport route serving the modelled area should be surveyed at least once. These should be surveyed over the whole period for which the model will apply. Information on bus routes and services can be found using the NPTDR database referred to in paragraph 2.6.9.

3.4.17 This survey method can only be conducted with the permission of the bus or train operating companies.

Face-to-face at-stop or at-station surveys

3.4.18 Face-to-face interviews with passengers at stops and stations can provide the same scope and quality of information as on-board interviews. However, interviews are required at all stops and stations for a complete picture of travel in the modelled area. This is often uneconomic and so surveys of this kind are best targeted at major generators or attractors of travellers which, in practice, means town and city centres. It is important that all stops and stations in the targeted areas are included in the survey. If the operators will not allow surveys to be conducted on board their vehicles, the modeller will have no option but to rely on a combination of at-stop or at-station surveys and electronic ticket machine data.

3.4.19 While interviews can, in principle, be conducted with boarders and alighters at bus and light rail stops or train stations, people alighting are likely to be much less willing to be interviewed than those waiting for a service to arrive.

3.4.20 As with on-board face-to-face surveys, greater proportions of inter-peak travellers may be interviewed than peak travellers. The important consideration is that the sample size during both the peak and inter-peak periods should be sufficiently large to support the creation of acceptably reliable trip matrices for these periods.

3.4.21 Unlike on-board surveys, issues of long and short journey biases do not apply in that those making short hops are as likely to be approached for interview as those making longer journeys.

3.4.22 However, the data collected by this method are likely to be biased in that those passengers that arrive at the stop or station immediately before the vehicle arrives (‘runners’) will be under-represented. Where there are infrequent regular services, such as on trains, then this can lead to biases between regular and irregular passengers. Regular users of the service may time their arrival to the station more closely to the departure time. This bias may not be as prevalent on regular services where departure times cannot be easily predicted, such as for most urban bus routes.

3.4.23 As with on-board surveys, data collection should be conducted for the entire period for which the transport model will apply.

3.4.24 Whereas permission is often not required for surveys at bus stops, it will be required for platform or station entry or exit surveys.

Self-completion on-board surveys

3.4.25 Self-completion surveys should be regarded as a means of last resort. While, in principle, the scope of the information which may be gathered by this means is as comprehensive as that obtained by face-to-face interview, the quality is likely to be poorer in a number of respects. First, some information may not be provided. Secondly, some information may be inaccurate if the questionnaire is completed some time after the journey was made. And, thirdly, the sample of respondents will be self-selected and may therefore not be adequately representative of all travellers.

3.4.26 The method involves the distribution of self-completion questionnaires to passengers as they board vehicles. Often these questionnaires would be collected from passengers as they alight from the vehicle.

3.4.27 Although forms should be distributed with reply-paid envelopes, in order to maximise response rates, it is recommended that measures be taken to ensure that forms are completed during the journey and returned to the surveyors when alighting. For this reason survey forms should be very brief and uncomplicated. Pens and pencils should be distributed to boarders along with the survey forms.

3.4.28 The main advantage of the self-completion survey compared with face-to-face interviewing is that all passengers have an equal likelihood of being given a survey form, regardless of whether they are making long or short journeys and whether travelling during peak or inter-peak times. However, return rates do vary due to self-selection biases.

3.4.29 Where self-completion surveys are collected from, as well as distributed to, passengers, return rates of forms of up to 95% can be achieved. However, large proportions of these forms will be incomplete. Response rates to individual questions in bus surveys can vary from between 40% and 70%. Response rates to questions may be higher on trains.

3.4.30 The same issues regarding which routes and times of day to survey apply to the self-completion on-board survey as to the face-to-face interview on-board survey. Similarly, permission is necessary from operating companies for surveyors to work on board vehicles.

Self-completion at-stop and at-station surveys

3.4.31 This method is conducted by distributing self-completion forms to passengers as they enter or exit stations or as they wait at or alight at bus stops.

3.4.32 The main advantage of this method compared with an at-stop face-to-face interview is that more passengers are likely to accept survey forms than would be willing to participate in an interview. This will apply to both those who arrive at stops or stations just before the train/bus arrives and also to those on time-critical journeys. Bus passengers waiting at stops would be more likely to accept a survey form even if their bus is approaching or already at the bus stop.

3.4.33 As with on-board surveys, at stop and at station surveys will suffer from inaccuracies and limitations to the data that can be collected. Depending on the survey materials, return rates of distributed questionnaires could be as low as 15% to 25%. Furthermore, some questions will not be answered.

3.4.34 Again, sampling biases may occur depending on which stations and stops are surveyed and what times of day are covered. Litter generation of discarded survey forms can be a significant problem. Permission must be sought from station managers to distribute survey forms within stations.

Summary of the preferred approach

3.4.35 Table 2 sets out the main advantages of each survey method, along with biases that might occur, the practical difficulties of the method and likely response rates. Details of when each different type of method is most appropriate are also shown.

3.4.36 In summary, the best approach is for an adequate sample of face-to-face interviews to be conducted on-board a sample of public transport vehicles on each service in the modelled area. However, operators may not allow interviews to be conducted on their vehicles.

3.4.37 The next best approach is to conduct face-to-face interviews at stops and stations. While comprehensive coverage of Network Rail stations may be affordable, it is likely that, to be cost-effective, surveys at bus stops would have to be confined to particular areas. This means that some other source of data will be required to supplement the at-stop surveys.

3.4.38 Electronic Ticket Machine (ETM) data could be useful in validating on-board surveys and supplementing at-stop surveys. However, not all operators are willing to make their information available, especially in circumstances where the proposed scheme would compete with the operator's existing services. In these instances, self-completion questionnaires issued at stops may be used in combination with face-to-face interview at-stop surveys. In such an approach, face-to-face interviews would be concentrated in the town or city centre and self-completion questionnaires distributed elsewhere. Both techniques should be used at a small sample of stops so that a check for bias can be made.

Table 2 Advantages and Disadvantages of Different Public Transport Survey Methods

Method	Advantage	Potential Bias	Practical Difficulties	Response Rates	Appropriate Use
On-board face-to-face survey	Individual interviews of high quality	Fewer short-journey passengers, fewer peak passengers	Impossible to administer on crowded vehicles	From 10% on crowded vehicles to 90% on quiet services	Not busy buses, particularly in peak times.
At bus stop surveys; on-platform surveys	Individual interviews of high quality	Fewer 'runners', fewer peak passengers, no passengers who use those stops not surveyed	Unproductive, expensive if administered at all bus stops	From 5% at peak times to 60% inter-peak of those visiting survey bus stops	When there is no permission for on-board interviews, where it is acceptable that some stops are not surveyed. Where attitudes towards stop facilities required.
Entry/exit surveys at stations	Individual interviews of high quality	Fewer time-critical journeys, fewer 'runners', fewer peak passengers.	Very difficult to recruit at high passenger flows	From very few at peaks to 90% inter-peak	Not during morning peak hours at busy stations.
On-board self completion survey	Highly productive	Self-selection biases. Some response bias against short-hop passengers	Poor quality and completeness of data, limited scope of data.	Up to 95% of survey forms returned but as low as 40% of questions answered	On highly crowded services, where only limited data required.
At bus stop or station self-completion survey	Productive	Self-selection biases, fewer time-critical journeys	Difficult to cover all bus stops on network, Poor quality and completeness of data, limited scope of data	Between 15% and 25% returned, with about 70% of questions answered	Where no permission for on-board survey, only limited data required, where it is acceptable that some stops are not surveyed.

4 References

Highways England *Design Manual for Roads and Bridges (DMRB) available at* *http://www.dft.gov.uk/ha/standards/dmrb/index.htm*

5 Document Provenance

This unit consists of restructured and edited material from the following guidance units that existed in the previous WebTAG structure and from DMRB at August 2012

Data Sources (WebTAG Unit 3.1.5), the primary WebTAG unit containing guidance on data sources, dated June 2003.

Scope of the Model (WebTAG Unit 3.10.2), has been incorporated where relevant to discuss principles that are applicable to demand modelling, dated April 2011.

Road Traffic and Public Transport Assignment Modelling (WebTAG Unit 3.11.2), contains principles of data collection for assignment modelling, dated January 2006.

Highway Assignment Modelling (WebTAG Unit 3.19), has been incorporated where relevant to discuss principles that are applicable to highway assignment modelling, dated August 2012.

DMRB Volume 5 Section 1 Part 4: TA 11/09 Traffic Surveys by Roadside Interview

DMRB Volume 12 Section 1 Part 1: Traffic Appraisal Manual

DMRB Volume 12 Section 2 Part 1: Traffic Appraisal in Urban Areas